

## ЛЕКЦИЯ №7. РЕГРЕССИОННЫЙ АНАЛИЗ

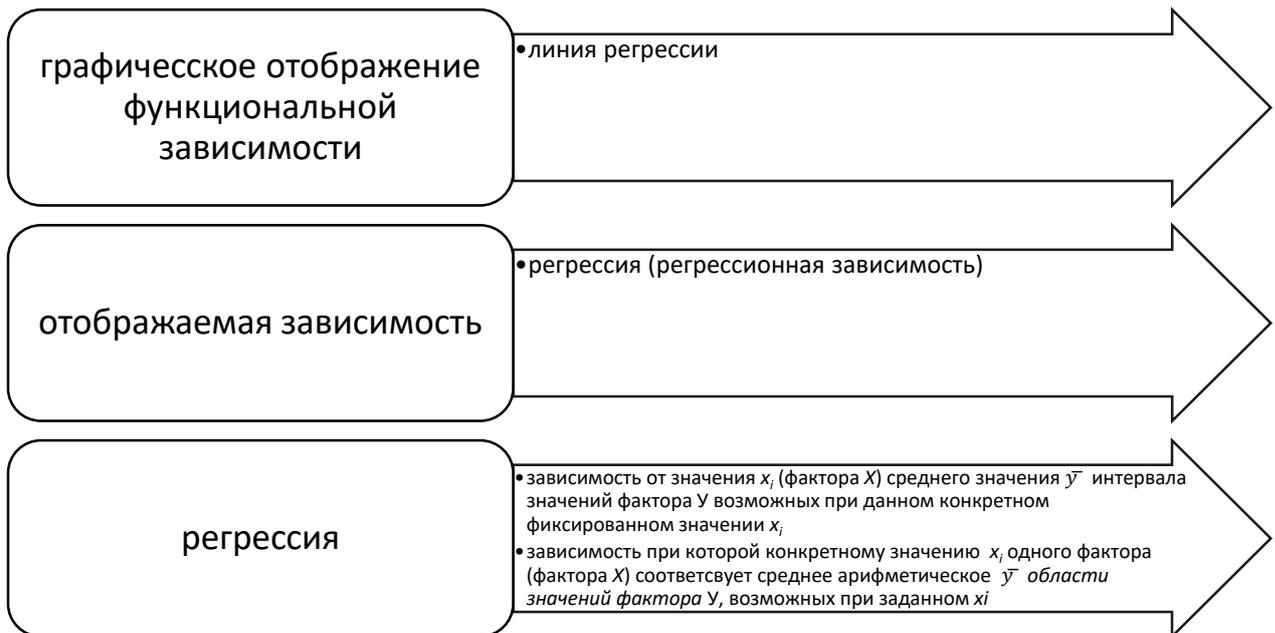
### План:

1. Регрессия.
2. Прогнозирование мировых достижений.
3. Прогнозирование спортивной одаренности.

**Ключевые слова:** регрессионный анализ, регрессионные модели, параметры уравнения регрессии, уравнения регрессии, прогнозирование мировых достижений, метод экстраполяция, спортивное прогнозирование, спортивная одаренность, способности, задатки, наследственное влияние, дефенитивные признаки, ювенильные признаки.

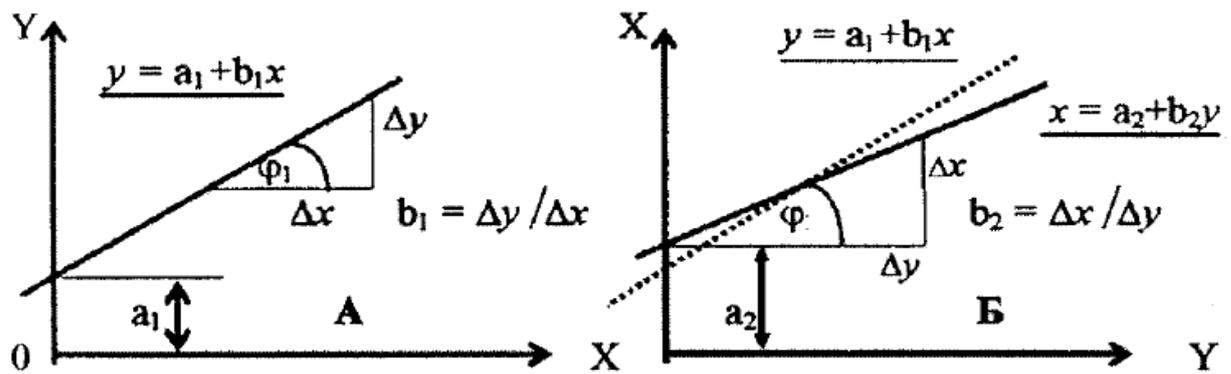
### Регрессия.

Корреляционное поле графически отображает статистическую зависимость, при которой каждому конкретному значению одного фактора соответствует интервал значений. Если же каждому конкретному значению фактора противопоставить среднее значение по интервалу, на графике получается ряд точек, лежащих на некоторой линии, то есть графическое отображение функциональной зависимости: каждому конкретному значению одного фактора в таком случае соответствует тоже одно конкретное значение.



Это по сути статистическая зависимость, отображенная в форме функциональной.

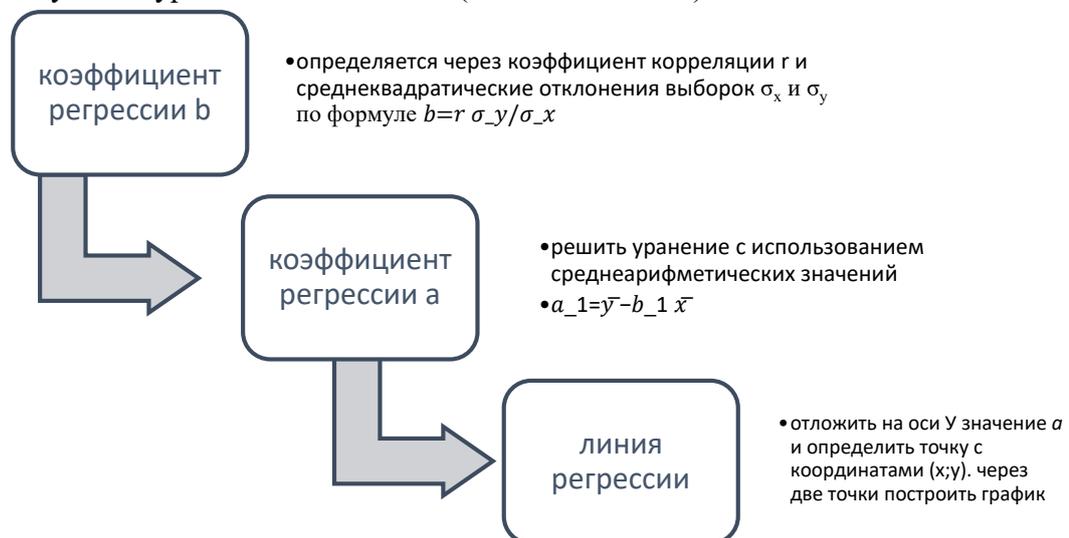
Таким образом, по существу переход от корреляционной зависимости к регрессионной (от корреляции к регрессии) - это формальный переход от статистической зависимости к функциональной, но при этом сохраняется ясное понимание того, что функциональной зависимостью здесь определяется лишь ориентир, вблизи которого с большой вероятностью и ожидаются значения интересующего нас зависимого фактора. Например, если взять зависимость «должной» массы тела от роста, дается ориентир: например, при росте 170 см «табличная» («должная») масса тела - 65кг. Он рассматривается как некий усредненный ориентир, и в зависимости от типа телосложения от этой по самому своему замыслу не обязательной, а очень условной нормы можно в довольно широких пределах отклоняться.

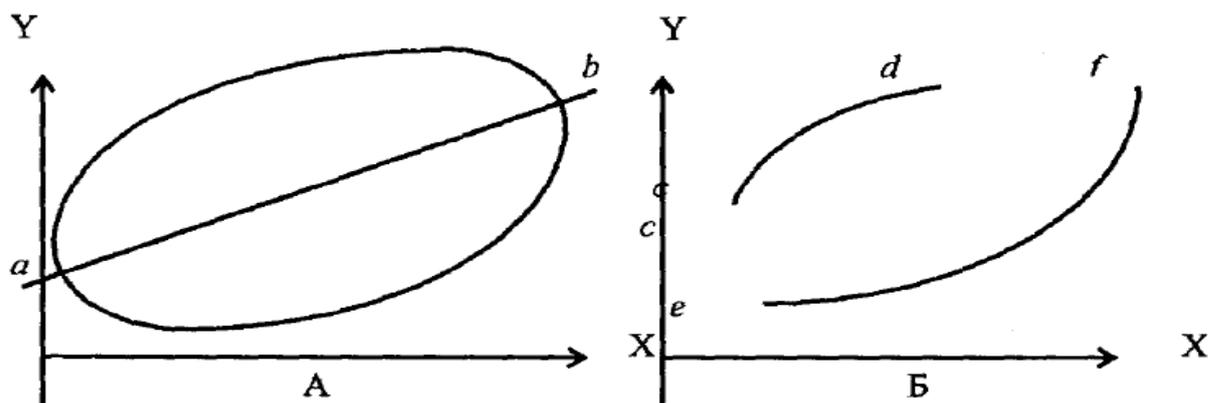


Линии прямой (А), прямой и обратной линейной(Б) регрессии

На рисунке (А) показано графическое отображение регрессии, соответствующей уравнению  $Y = a + bx$  — уравнению линейной регрессии, на графике это прямая. Приведенное уравнение представляет собой уравнение прямой, в котором коэффициент (параметр)  $a_1$  - отрезок, отсекаемый этой прямой от оси абсцисс, он может быть положительным или отрицательным (т.е. может лежать выше или ниже оси абсцисс), коэффициент (параметр)  $b_1$  - так называемый угловой коэффициент прямой, равный тангенсу угла между ней и осью абсцисс, то есть,  $b_1 = \Delta y / \Delta x$ . Коэффициент уравнения регрессии  $b_1$ , определяющий основную часть связи между факторами, называют коэффициентом регрессии. Его изменение влечет за собой изменение угла наклона  $\varphi$  линии регрессии, тогда как изменение величины коэффициента  $a$ , не меняя угла  $\varphi$ , ведет к ее смещению вверх или вниз параллельно самой себе на величину изменения  $a$  (то есть на  $\Delta a$ ).

На рисунке (Б) показаны линии прямой регрессии, соответствующей уравнению  $Y = a_1 + b_1 X$  (пунктирная линия) - той же, что и на рисунке (А), и обратной регрессии, соответствующей уравнению  $X = a_2 + b_2 Y$  (сплошная линия).





*Линии регрессии: a - отображение («замена») корреляционного поля линией линейной регрессии ab, б — линии 2 разных нелинейных регрессий*

Линию линейной регрессии, определенным образом отображающую корреляцию, можно «на глаз» провести в корреляционном поле, если его форма близка к эллиптической (т.е. корреляцию можно считать линейной) как длинную ось эллипса, называя «линией свободной руки». Более сложный тип регрессии - нелинейная регрессия - графически отображается некоторой кривой.

Регрессия может быть представлена и таблично: так, например, все таблицы расчета очков, соответствующих конкретным результатам в легкоатлетических многоборьях, построены на базе соответствующих регрессий, на базе регрессий построены и ростовесовые таблицы, различные таблицы норм, многие математические таблицы.

## **I Прогнозирование высших мировых достижений**

Прогнозирование высших мировых достижений (как правило, не более чем на 1-2 олимпийских цикла) необходимо для планирования подготовки спортсменов, а также для определения модельных характеристик (прежде всего, показателей физической и технической подготовленности), соответствующих предполагаемым результатам. Понятно, что на динамику роста спортивных достижений влияет много факторов, и в этой ситуации невозможно точно предугадать, какой именно результат будет показан на ближайших ответственных соревнованиях (чемпионат мира, Европы, Олимпийские игры). Такой прогноз может быть только приближенным. Накопленный опыт прогнозирования, однако, показывает, что точность прогноза на ближайший олимпийский цикл удовлетворительна: как и полагается в случае нормального распределения, примерно 2/3 всех результатов находятся в пределах одной стандартной ошибки от наиболее вероятного прогнозируемого значения.

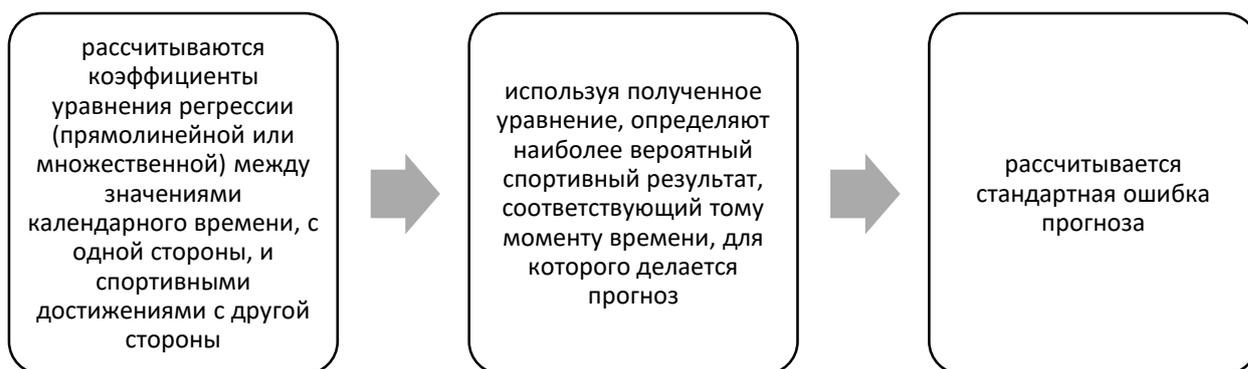
В видах спорта с объективно измеряемыми результатами для прогнозирования используют метод регрессионного анализа. Так, по соответствующему уравнению регрессии определяют наиболее вероятный спортивный результат. Рассчитывается также стандартная ошибка этого прогноза.

Данный подход заключается в том, что весь процесс роста результатов  $y(t)$  представляют в виде суммы:

$$y(t) = x(t) + z(t)$$

где  $x(t)$  – неслучайная составляющая (временной тренд),

$z(t)$  – случайная функция времени, «шум», наложенный на закономерный процесс роста результатов.



Подобная операция определения по ряду известных значений величины других её значений, находящихся за пределами этого ряда, в науке называется экстраполяцией. При прогнозе высших мировых достижений экстраполяция равносильна предположению, что эти достижения будут расти в будущем примерно так же, как они росли в прошлом. Точность экстраполяции тем больше, чем короче период, на который она делается, и чем продолжительнее история вида спорта.

Если в какой-то спортивной дисциплине появляется значительное новшество (резкое изменение методики тренировки или спортивного инвентаря), то спортивные достижения начинают расти быстрее и точность экстраполяции падает. При подобных скачках результатов прогноз осуществляют, комбинируя методы экстраполяции с методом экспертных оценок.

## Спортивное прогнозирование

- одна из важных областей применения регрессионного анализа в спортивных исследованиях

Другая важная область применения регрессионного анализа в спортивных исследованиях связана с прогнозированием, но в несколько другом понимании этого термина. Очень часто предметом исследования является такой признак, который непосредственно измерить затруднительно или невозможно. В то же время известно, что изучаемый признак связан с другими признаками, которые измеряются сравнительно просто. Тогда пытаются подобрать модель предполагаемой зависимости и по этой модели прогнозировать значения неизмеряемого зависимого признака, основываясь на значениях других, легко измеряемых признаков. Прогнозируемые таким образом значения неизмеряемых признаков называются в статистике предикторами. В этом случае также используются регрессионные модели, потому что оцениваемая величина является случайной: помимо контролируемых факторов, влияние которых учитывается значениями измеряемых признаков, она зависит и от множества других случайных факторов, которые контролировать не удается.

Модель простой линейной регрессии, отражающая зависимость значений  $Y_i$  зависимой величины  $Y$  от значений  $x_i$  независимой переменной  $x$  в генеральной совокупности, описывается уравнением:

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

В этом уравнении  $\alpha$  и  $\beta$  – неизвестные параметры уравнения регрессии, а  $\varepsilon_i$  – случайные ошибки, представляющие собой случайные отклонения значений  $Y_i$  от линии регрессии:

$$\varepsilon_i = Y_i - (\alpha + \beta x_i)$$

Истинное уравнение регрессии  $m_{y/x} = \alpha + \beta x$  обычно неизвестно, потому что не имеется возможности наблюдать всю генеральную совокупность. Единственное, что можно сделать, чтобы построить линию регрессии, - это провести выборочное исследование и по экспериментальным данным оценить генеральные параметры  $\alpha$  и  $\beta$ . Пусть получена выборка объема  $n$  наблюдений  $y_i$  зависимой случайной величины  $Y$ , соответствующих значениям  $x_i$  независимой переменной  $x$ .

Оценки параметров  $\alpha$  и  $\beta$ , которые получаются по выборочным данным, обозначаются соответственно  $a$  и  $b$ . Для определения оценок  $a$  и  $b$  чаще всего применяется метод наименьших квадратов. Суть этого метода в том, что отыскиваются такие значения  $a$  и  $b$ , которые обеспечивают минимум суммы квадратов отклонений измеренных значений  $y_i$  от прямой линии, задаваемой параметрами  $a$  и  $b$ , т.е.

$$\min_{a,b} \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

Таким образом, по методу наименьших квадратов получаем эмпирическое уравнение некоторой прямой:

$$Y_x = a + bx.$$

Здесь  $Y_x$  - принятое обозначение для оценки величины  $Y$  при заданном значении  $x$ .

Значения  $a$  и  $b$  по методу наименьших квадратов находятся из решения системы так называемых нормальных уравнений:

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

$$a = \bar{y} - bx$$

где  $\bar{x}$  и  $\bar{y}$  - выборочные средние арифметические.

Обычно  $b$  называют коэффициентом регрессии, а – свободным членом уравнения регрессии.

Мерой качества приближенного описания реальной зависимости между величинами  $Y$  и  $x$  с помощью уравнения линейной регрессии является стандартное отклонение значений  $y_i$  от регрессионной прямой, вычисляемое по формуле:

$$S_{yx} = \sqrt{\frac{\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i}{n - 2}}$$

$S_{yx}$  является мерой точности предсказания значений случайной величины  $Y$  по заданным значениям величины  $x$ , поэтому  $S_{yx}$  называют также стандартной ошибкой предсказания.

Если в результате проведенной проверки нет оснований сомневаться в адекватности линейной модели, то необходимо проверить гипотезу о том, что в действительности в генеральной совокупности отсутствует линейная регрессия, а то, что полученный коэффициент регрессии  $b$  отличен от нуля, объясняется только случайностью выборки.

Гипотеза  $H_0$  проверяется с помощью стандартного  $t$ -критерия Стьюдента, определяемого по формуле:

$$t = \frac{|b|}{\frac{S_{yx}}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}}$$

где  $|b|$  - абсолютная величина коэффициента регрессии,

$S_{yx}$  - стандартная ошибка предсказания.

Если значение критерия  $t \geq t_\alpha$ , то нулевая гипотеза отклоняется, и можно сделать вывод, что линейная регрессия значима на уровне значимости  $\alpha$ . В противном случае гипотеза  $H_0$  принимается.

Существует возможность применения регрессионного анализа для прогнозирования спортивного результата по данным тестирования функционального состояния организма спортсмена в разные периоды изменчивости его физиологических функций. В случае невозможности применения сложных инструментальных методов измерения ( $VO_2$ , лактата и пр.), можно прибегать к более простым.

На начальном этапе прогноза следует убедиться в наличии корреляции результатов тестирования с критерием информативности (спортивный результат), а лишь затем осуществлять расчёты по выведению уравнений линейной регрессии.

После подобного установления высоких корреляционных отношений следует переходить к расчёту уравнения линейной регрессии:

$$y = a + b \cdot x,$$

где  $y$  – искомое значение (прогнозируемый спортивный результат);

$a$  – свободный член;

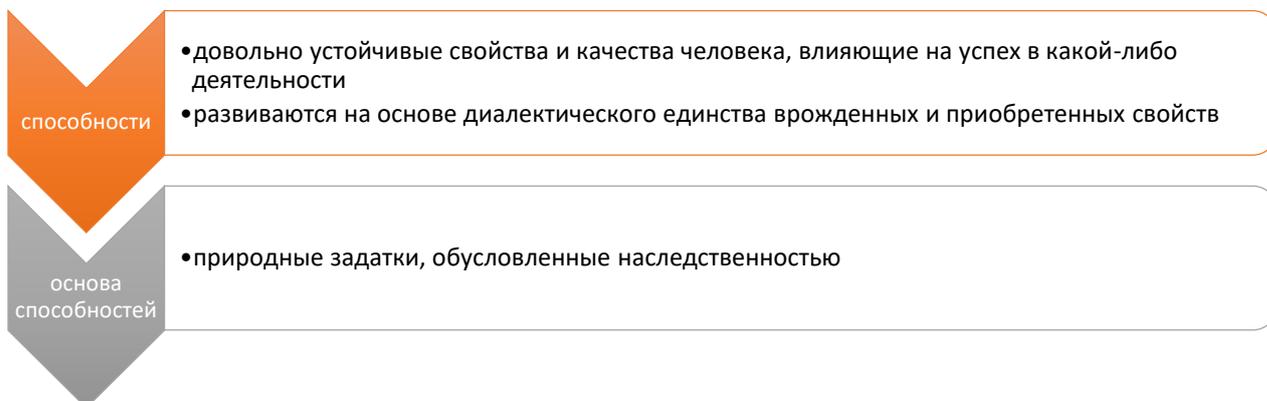
$b$  – коэффициент регрессии;  $x$  – заданное значение (численная величина в единицах измерения в тесте – ЧСС,  $VO_2$ , концентрация лактата, путь, время или скорость прохождения отрезков дистанции и т. п.).

Расчёт и построение линии (линейной или криволинейной) регрессии и её стандартной ошибки позволяют количественно определить степень влияния заданного значения на искомое.

## II Прогнозирование спортивной одаренности

Особую прогностическую ценность имеет спортивная одарённость. Она характеризуется определённым сочетанием двигательных и психологических способностей, а также анатомо-физиологических задатков, создающих в комплексе потенциальную возможность для достижения высоких спортивных результатов в

конкретном виде спорта. От одарённости зависит не сам успех в спорте, а только возможность его достижения.



Основой способностей являются природные задатки (или потенциальные способности), обусловленные наследственностью. Однако упорный труд может компенсировать отсутствие или слабость способностей. Но если сравнивать упорный труд нескольких спортсменов, то более способный будет иметь преимущество.

Иногда вместо слова «задатки» используют термин «потенциальные способности»; в этом случае способности, проявляемые в данный момент времени (регистрируемые, например, с помощью тестов) называют актуальными способностями.

Прогноз спортивной одарённости может быть сделан на основе изучения либо стабильности показателей, либо наследственных влияний.

Стабильность двигательных качеств определяется по темпам их прироста у ребёнка в течение не менее полутора лет, а при взрослении – 1, 2 раза в год.

Важно отметить, что в период полового созревания (у девочек 11-14 лет, у мальчиков 13-15 лет) стабильность снижается; в этом возрасте прогноз малоэффективен. К примеру, прогнозировать вес тела у взрослого человека по данным в детские годы затруднительно.

При изучении стабильности показателей вопрос ставят так: насколько стабильны характеристики ребенка в процессе его развития? Например, при поступлении в первый класс у детей измерили ряд показателей (рост, результат в беге на 30 м и т.д.). Будут ли дети, занявшие первые места в беге также лидировать через 9 лет, по окончании школы? Будут ли самые высокие также оставаться выше всех, а маленькие по-прежнему оставаться самыми низкими?

Значение признака в детские годы называют ювенильным, в конце наблюдаемого периода – дефинитивными. Можно ли по ювенильным значениям прогнозировать дефинитивные?

Если такие наблюдения проводились систематически (например, раз в год), то можно отобразить их графически. Такие графики называют физиограммами. Чтобы оценить стабильность какого-либо показателя у всей группы испытуемых, рассчитывают коэффициенты корреляции между ювенильными и дефинитивными значениями признака. Такой коэффициент корреляции называют коэффициентом стабильности.

Результат каждого последующего тестирования можно рассматривать как сумму результата предыдущего тестирования и величин прироста показателя:

$$x_{t+1} = x_t + \Delta x$$

где  $x$  и  $x_{t+1}$  – значения показателя в последовательные моменты времени  $t$  и  $t+1$ ;

$\Delta x$  – прирост показателя за данный период.

При этом корреляция ювенильного и дефинитивного признаков – это корреляция между  $x_t$  и  $(x_t + \Delta x)$ . Из теории корреляции известно, что в таком случае все определяет  $\Delta x$ , а именно его взаимосвязь (корреляция) с  $x_t$ . Если такой корреляции нет (т.е. величина прироста не зависит от исходного уровня; у высоких и низких, сильных и слабых результаты растут одинаково), то корреляция ювенильных и дефинитивных значений равна просто отношению их стандартных отклонений:

$$r_{t(t+1)} = \frac{\sigma_t}{\sigma_{t+1}}$$

Когда  $\Delta x$  не коррелирует с  $x_t$ , прогноз по ювенильным значениям оказывается довольно точным. К сожалению, это бывает не всегда. Что касается двигательных возможностей человека, то здесь нередко отрицательные корреляции: прирост результатов тем меньше, чем выше их исходный уровень. В таких случаях прогноз дефинитивных значений по ювенильным оказывается невозможным; их следует прогнозировать по темпам прироста.  $X_{t+1}$  всегда будет коррелировать либо с  $x_t$ , либо с  $\Delta x$ , либо и с тем и с другим вместе. Опыт показывает, что для удовлетворительного прогноза следует наблюдать ребенка в течение не менее полутора лет.

Таким образом, для прогнозирования спортивной одаренности детей на основе изучения стабильности показателей следует опираться на:

- 1) Коэффициент стабильности;
- 2) Корреляцию дефинитивных признаков с темпами прироста показателей в течение не менее полутора лет.

Изучение наследственных влияний показывает, что ими в большей или меньшей степени определяются многие анатомо-физиологические показатели, от которых зависят успехи в ряде видов спорта. Основным методом изучения наследственных влияний является исследование близнецов. Близнецы бывают однояйцевые (монозиготные, развившиеся из одной оплодотворенной яйцеклетки) и двуяйцевые (дизиготные, развившиеся из двух оплодотворенных яйцеклеток). Монозиготные близнецы наследственно тождественны. Дизиготные близнецы находятся между собой в такой же степени родства, как и обычные братья и сестры; они генетически различны. Совпадение (конкордантность) или несовпадение (дискордантность) каких-либо признаков у монозиготных и дизиготных близнецов служит основой для определения наследственно обусловленных свойств. Количественно степень наследственных влияний оценивается так называемым коэффициентом наследственности. Он может изменяться по величине от 0 до 1. При коэффициенте, равном нулю, наследственное влияние на данный признак отсутствует, а если этот коэффициент равен единице, признак полностью находится под генетическим контролем.

#### ***Вопросы для контроля:***

1. На какой срок, как правило, выполняется прогнозирование высших мировых достижений?
2. Для чего нужно прогнозирование высших мировых достижений?
3. Какой математический метод применяется для прогнозирования высших мировых достижений?
4. Что такое экстраполяция?
5. Что может повлиять на точность результатов прогноза?

6. Приведите уравнение простой линейной регрессии
7. Из каких показателей складывается спортивная одаренность?
8. Что представляют собой спортивные способности?
9. Что лежит в основе способностей?
10. На какие показатели следует опираться при прогнозировании спортивной одаренности детей?